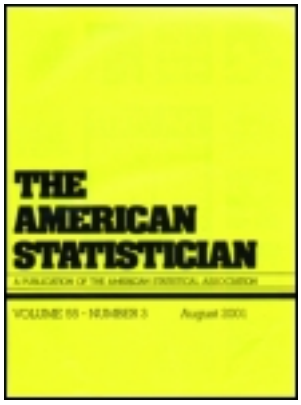

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Mean and Variance of Truncated Normal Distributions

Donald R. Barr & E. Todd Sherrill [a]
[a] U.S. Army, 304 Miller Loop, Fort Benning, GA, 31905, USA
Published online: 17 Feb 2012.

Mean and Variance of Truncated Normal Distributions

Donald R. BARR and E. Todd SHERRILL

Maximum likelihood estimators for the mean and variance of a truncated normal distribution, based on the entire sample from the original distribution, are developed. The estimators are compared with the sample mean and variance of the censored sample, considering only data remaining after truncation. The full- and censored-sample estimators are compared using simulation. It is seen that, surprisingly, the censored-sample estimators generally have smaller mean square error than have the full-sample estimators.

KEY WORDS: Censored-sample estimators; Full sample estimators.

1. INTRODUCTION

In a variety of applications the mean and variance of normal distributions that have been truncated in various ways are needed. For example one may be concerned with the population of remaining SAT scores for college admissions candidates, when all candidates with scores below a fixed threshold or "screening value" have been eliminated. If the original population is normally distributed, the "screened population" has a truncated normal distribution. This is the same as the conditional distribution of a randomly selected score, given it exceeded the threshold; this perspective suggests how one could simulate outcomes using a simple rejection rule with a normal generator. (More efficient generators are discussed by Law and Kelton 1991).

Hald (1952) and others gave the cumulative distribution function (CDF) and density of the truncated normal distribution; Kececioglu (1991) displayed a simple expression for the CDF when the truncation is expressed as a fraction of the original population. In the following, we derive the mean (in a form involving the normal CDF) and variance of truncated normal distributions, in terms of the parameters of the original normal distributions. Although the variance cannot be given in simple closed form, curiously it can be given in terms of a chi-square distribution with three degrees of freedom. Thus, armed with normal and chi-square CDF algorithms, such as those routinely implemented in spreadsheet software, numerical values of these moments can easily be calculated. Johnson and Kotz (1970) gave expressions for these moments, but their expression for the variance is lengthy and seems harder to implement than the one we give. Their expressions can be obtained from ours using the recursion (see Abramowitz and Stegun 1972)

$$C_3(t^2) = C_1(t^2) - \frac{te^{-t^2/2}}{\sqrt{\pi/2}},$$

where C_m is the chi-square with m degrees of freedom CDF, together with the fact that the square of a standard normal random variable is distributed as C_1. Fisher (1931) expressed these moments in terms of "Hh" functions, which also seems somewhat involved.

There is a long and rich body of literature relating to estimation of the parameters of the original population, based on data from the truncated distribution (i.e., censored data). Cohen contributed much in this area (1949, 1950, 1959, and 1961), along with Gupta (1952), Halperin (1952), and others. In particular, the material in Cohen (1991, chap. 2) may be useful to readers. Johnson and Kotz (1970) gave an excellent review of the literature on maximum likelihood estimation in various data-censoring situations.

However, in some applications we are concerned with the reverse: we have available the entire sample from the original population (or, in some cases, we may know the parameters of the original distribution), and we want to estimate (or compute) the mean and variance of the population after truncation. By the invariance principle of maximum likelihood estimators, MLEs of the mean and variance of the truncated distribution, based on the full dataset, can be obtained by substituting $\bar{x}$ and s^2 into the foregoing expressions. On the other hand, one could simply compute the sample mean and variance of the censored sample. We might expect the performance of the latter estimators to be inferior to that of the full-sample estimators, because they are based on fewer data. Surprisingly, it turns out that is not the case, as we show in the following.

2. EXAMPLES

Army Selection Boards. The Army uses centralized Army-wide selection boards to select officers for promotion and advanced military schooling. Each selection board determines a performance-based "order of merit" ranking of the officers under its consideration. Generally, officers under consideration for promotion or advanced schooling that are not selected for the new grades or schools either leave or are separated from the Army. For example, officers being considered for promotion to Lieutenant Colonel (LTC) have successfully passed five such selection boards. Data for recent years show the selection rates of these boards averages about 78%. Thus, very roughly, the fraction of officers remaining after five boards is about 30% of the original population ($.78^5 = .29$). (There are losses of officers for reasons other than non-selection, but non-selection, or the threat of non-selection, accounts for most of the decreases in the numbers of officers at the succeeding higher ranks.) If we assume the original population of officers has

Donald R. Barr is Retired, P.O. Box 2071, Paradise, CA 95967-2071. E. Todd Sherrill is a Major, U.S. Army, 304 Miller Loop, Fort Benning, GA 31905.

normally distributed "performance," and selection boards select officers with the highest performance, then a LTC selection board is effectively considering a truncated normal population of "performance," with a truncation point corresponding to the 70th percentile of the original normal population.

For a standard normal distribution, truncation at the 70th percentile would correspond to a truncation point, t, of .53. The variance of such a truncated normal is about .26 (see Figure 2). We conclude the variance of the population under consideration by the LTC board is only about one-fourth that of the original population. This relatively smaller variance makes it more difficult for the board to discriminate among the officers under consideration. Members of selection boards for the higher ranks are sometimes quoted as saying, "All the officers look about the same," and criticisms of the officer evaluation report system, upon which performance ranks are based, are frequently offered. But, the rapid decrease in variance as t increases goes a long way toward explaining this phenomenon. Even if the officer evaluation report system were perfect, and used optimally, there would inevitably be relatively little difference in performance scores at the higher ranks, due to the effects of truncation.

"Lost" Rounds in Artillery Adjustment. Truncation of a normal distribution above a certain value provides a useful model for artillery rounds fired at a target located on a hillside at a gunnery range. In many artillery applications it is reasonable to assume the lateral miss distances (called "deviations") and miss distances along the gun-target line (called "range misses") are independent random variables. The central limit theorem suggests, and data analysis confirms, the distributions of deviations and range misses are usually modeled well by normal distributions. For simplicity, let us consider only range misses. In attempting to hit the target in a precision mission, the gun crew adjusts the elevation of the gun after each round has been fired and its impact observed relative to the target. The amount of adjustment, stated in terms of the expected translation of the mean range of impact on the ground, is some fraction of the observed (estimated) range miss of the round. (Interestingly, the fractions must form a divergent series in order for the process to converge; the harmonic series is often used, so the nth adjustment is $1/n$ times the nth observed range miss.) A round fired at the target may be "lost" because it misses the hillside entirely (i.e., the round goes over the ridge), so the range miss cannot be determined. Thus, the sample of observed range misses form a censored sample from a normal distribution, where the truncation point corresponds to the range of the ridge. Sample moments computed from recorded range miss data are thus estimates of the corresponding moments of the truncated distribution, not the original normal distribution of miss distances.

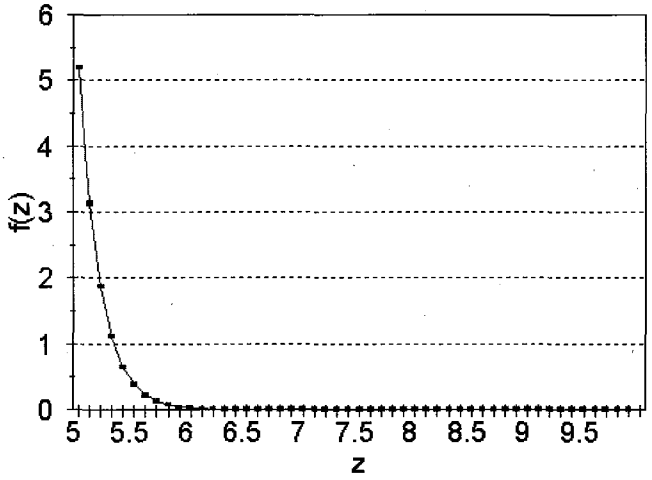

Figure 1. Plot of the density of a N(0,1) population truncated below at 5.0. The mean of this distribution is approximately 5.2 and the variance is approximately .03.

3. MEAN AND VARIANCE IN THE STANDARD NORMAL CASE

Let Z be a standard normal random variable truncated below at a fixed point, t. The density of Z is

$$f(z) = c(t)e^{-z^2/2}; \quad z \geq t,$$

(of course, $f(z)$ is zero for $z < t$), where $c(t) = 1/[\sqrt{2\pi}(1-\Phi(t))]$, and $\Phi(t)$ is the standard normal CDF. This density does not have the shape one might expect, especially for large truncation points. For large t, say $t > 5$, the probability density is large just to the right of t, and it decreases very rapidly toward zero as its argument increases (see Figure 1). Thus, for large t we would expect the mean of Z to be just above t and the variance of Z to be near zero. Johnson and Kotz (1970) showed several example density plots for various truncation schemes and assert, "When the degree of truncation is large, this distribution...[looks like] a rectangular or trapezoidal distribution." Such is not the case for the "one-sided" truncation we are considering.

The mean of Z is

$$E(Z) = c(t)\int_t^\infty ze^{-z^2/2}dz$$
$$= -c(t)e^{-z^2/2}|_t^\infty = c(t)e^{-t^2/2}. \quad (1)$$

One can easily see that when the t is small, $c(t) \approx 1/\sqrt{2\pi}$, so $E(Z) \approx \varphi(t)$, where φ is the standard normal density. Thus, when $t < -5$, $E(Z)$ differs little from 0, the mean of the original population, as one would expect. As t increases, $E(Z)$ increases, and for increasingly large truncation points, $E(Z)$ approaches t from above, again in accord with intuition.

To calculate

$$V(Z) = E(Z^2) - E^2(Z), \quad (2)$$

we concentrate on evaluating $E(Z^2)$. By definition, for $t \geq 0$,

$$E(Z^2) = c(t)\int_t^\infty z^2e^{-z^2/2}dz = \frac{c(t)}{2}\int_t^\infty ze^{-z^2/2}2zdz$$
$$= \frac{c(t)}{2}\sqrt{2\pi}\int_{t^2}^\infty \frac{1}{2^{3/2}\Gamma(3/2)}u^{1/2}e^{-u/2}du,$$

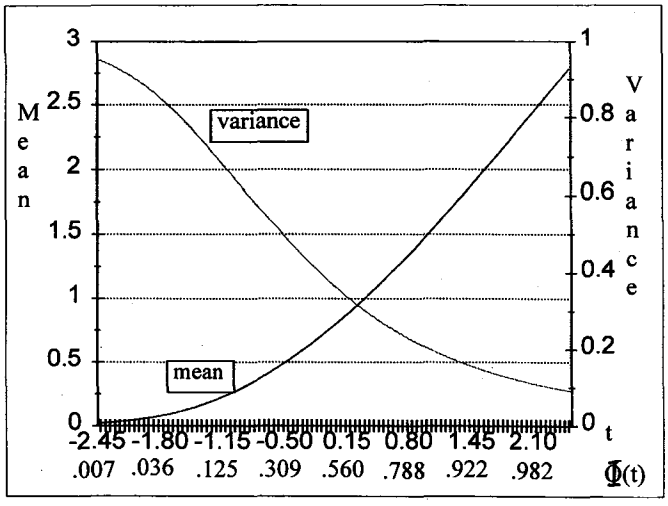

Figure 2. Plot of the mean and variance of a truncated standard normal distribution, as a function of the truncation point, t, and the fraction truncated, $\Phi(t)$.

where we made the substitution $u = z^2$. The third integral can be recognized as the integral of a chi-square density function with three degrees of freedom, so it is equal to $1 - C_3(t^2)$. It follows that, for $t \geq 0$,

$$E(Z^2) = c(t)\sqrt{\frac{\pi}{2}}[1 - C_3(t^2)]. \tag{3}$$

(As a check, at $t = 0$ this gives

$$E(Z^2) = \frac{1}{\sqrt{2\pi} \cdot 1/2} \cdot \sqrt{\frac{\pi}{2}} \cdot (1 - 0) = 1,$$

which should be the case because for a standard normal random variable X, $E(X^2|X \geq 0) = E(X^2) = V(X) = 1$ by a symmetry argument.)

The case for $t < 0$ requires a slight modification in the substitution $u = z^2$, because this transformation is not one-to-one. In this case integrate over the two branches $(t, 0)$ and $(0, \infty)$ to obtain

$$E(Z^2) = c(t)\sqrt{\pi/2}[1 + C_3(t^2)]. \tag{4}$$

Finally, using Equations (1)–(4),

$$V(Z) = c(t)\left[\sqrt{\frac{\pi}{2}}[1 \mp C_3(t^2)] - c(t)e^{-t^2}\right], \tag{5}$$

where "+" is used when $t < 0$ and "−" is used otherwise.

Figure 2 shows plots of $E(Z)$ and $V(Z)$ as functions of t, and proportions truncated, $\Phi(t)$. The figure was prepared by implementing expressions (1) and (5) in a spreadsheet. Note the rapid decrease in variance as t increases.

3.1 Computation for Nonstandard Normal Populations

The previous expressions can be used for normal distributions other than the standard normal. If X is a $N(\mu, \sigma^2)$ random variable truncated below at $t\sigma + \mu$, then $(X - \mu)/\sigma = Z$ is a standard normal random variable truncated below at t. Note

$$E(X) = \sigma E(Z) + \mu = \sigma c(t)e^{-t^2/2} + \mu, \tag{6}$$

where $c(t)$ is given above, and

$$V(X) = V(\sigma Z + \mu) = \sigma^2 V(Z), \tag{7}$$

where $V(Z)$ can be gotten from expression (5). Thus, the mean and variance in the general normal case is easily gotten from the standard normal case: for a $N(\mu, \sigma^2)$ variable truncated below at b, set $t\sigma + \mu = b$, solve for t, $t = (b - \mu)/\sigma$ and plug this value into expressions (6) and (7) using (5). For example, suppose X is a $N(1, 4)$ variable truncated below at 2. Then $t = 1/2$, so, using expressions (1) and (5), $E(Z) = 1.141078$ and $V(Z) = .26848$. Finally, using (6) and (7) we obtain $E(X) = 2E(Z) + 1 = 3.282156$ and $V(X) = 4V(Z) = 1.07392$.

3.2 Extension to Truncation Above

If X is a $N(\mu, \sigma^2)$ random variable truncated above at $-t\sigma + \mu$, then $(-X + \mu)/\sigma = Z$ is a standard normal random variable truncated below at t. Then $X = -\sigma Z + \mu$ so

$$E(X) = -\sigma E(Z) + \mu = -\sigma c(t)e^{-t^2/2} + \mu, \tag{8}$$

and

$$V(X) = \sigma^2 V(Z).$$

Therefore for a $N(\mu, \sigma^2)$ random variable truncated above at a, set $a = -t\sigma + \mu$ and solve for t, $t = (-a + \mu)/\sigma$, and plug into equations (8) using (5).

A similar line of reasoning can be used for other truncation schemes.

4. ESTIMATION

Given a random sample from an original $N(\mu, \sigma^2)$ population, where μ and σ^2 are unknown, how should one estimate the mean and variance of the truncated distribution? By the invariance property of maximum likelihood estimators, MLEs of the truncated mean and variance can be obtained by replacing μ and σ^2 in the foregoing expressions with the corresponding full-sample estimates $\overline{x}$ and $(n - 1)s^2/n$. A simple alternative would be to throw away all sample observations below t, consider the data that remain to be a random sample from the truncated population, and compute the sample mean and variance directly using the censored data (assuming at least two observations remain).

We consider two statistical issues associated with estimating moments of truncated normal distributions:

(a) How much better are the truncated mean and variance estimated by the full-sample MLEs, compared to the sample mean and variance of the censored sample?

(b) What if t is unknown (so we know only that a certain, observed number of the largest values in the sample are retained when the sample is censored)?

Issue (b) is probably not encountered often in actual applications, so we only comment on it briefly. It seems reasonable to use a naive estimator for t based on the relative frequency estimate of $p = P[X < t] = \Phi[(t - \mu)/\sigma]$. This suggests estimating t using $\hat{t} = s \cdot \Phi^{-1}(\hat{p}) + \overline{x}$. This estimated value could then be used in place of t if one were using the full-sample estimators described earlier.

Table 1. Estimates of Bias and MSE for Two Estimators of the Mean and Two Estimators of the Variance of Truncated Normal Populations

			Mean				Variance				
			MLE		Censored		MLE		Censored		
$t;\mu;\sigma^2-$	$P[X>t]$	n	bias	MSE	bias	MSE	bias	MSE	bias	MSE	P
0.00;	0.5	10	−.02	.14	−.01	.09	.00	.03	.00	.10	.011
.79788;		20	−.01	.07	.00	.04	.00	.01	.01	.05	.000
.36338		36	−.01	.04	.00	.02	.00	.01	.00	.02	.000
		50	−.01	.03	.00	.01	.00	.01	.00	.02	.000
		100	.00	.01	.00	.01	.00	.00	.00	.01	.000
0.50;	.31	10	−.03	.17	−.08	.13	.00	.02	−.01	.09	.136
1.14107;		20	−.01	.08	.00	.05	.00	.01	.00	.05	.006
.26848		36	−.01	.05	.00	.03	.00	.00	.00	.03	.000
		50	.00	.03	.00	.02	.00	.00	.00	.02	.000
		100	.00	.02	.00	.01	.00	.00	.00	.01	.000
1.00;	.16	10	−.04	.22	−.18	.17	.00	.01	−.02	.05	.513
1.52513;		20	−.02	.11	−.06	.09	.00	.00	−.01	.06	.151
.199098		36	−.01	.06	−.01	.04	.00	.00	.00	.04	.018
		50	−.01	.04	.00	.03	.00	.00	.00	.02	.002
		100	.00	.02	.00	.01	.00	.00	.00	.01	.000
2.00;	.02	10	−.06	.42	−.09	.38	.00	.00	.00	.00	.979
2.37321;		20	−.03	.19	−.07	.16	.00	.00	.00	.00	.925
.114279		36	−.02	.11	−.08	.08	.00	.00	−.01	.01	.811
		50	−.01	.08	−.08	.06	.00	.00	−.01	.01	.685
		100	−.01	.04	−.05	.04	.00	.00	.00	.02	.333

Notation is as follows:
t = truncation point;
μ = true mean of the truncated $N(0,1)$ population;
σ^2 = true variance of the truncated population;
n = sample size;
bias = (sample mean of estimator) − (true parameter value);
MSE = (sample variance of estimator) + (bias)^2; and
P = probability a sample of size n will have fewer than two observations greater than t.

We investigated Issue (a) using simulation; a summary to the results is given in Table 1. We generated 10,000 random samples from the standard normal distribution having the sample sizes shown in the rows of Table 1. We calculated the mean and variance of each sample and used equations (6) and (7) to estimate the mean and variance of the truncated normal, for the truncation points shown in the table. We also censored each of these samples at the given t values and computed the censored sample mean and variance, when enough data remained. As can be seen in the rightmost column of the table, in significant fractions of samples with large t and small n, insufficient data for computing the sample moments remained after censoring. This presents a dilemma. Simply dropping such cases leads to a comparison of estimators based on different sets of samples from the parent population; one is conditional, the other is not. One could eliminate the offending samples from the full-sample computations as well, but that would be impractical in applications. Therefore, we decided to modify the censored sample estimators as follows: use the censored sample moments when two or more values remain after censoring; use the MLEs based on the full sample otherwise. The properties of "censored sample" estimators shown in Table 1 are results obtained with this compromise. For cases with small P in the right column of Table 1, these estimators are essentially "pure" censored sample estimators, and it is especially interesting to compare them with the corresponding MLEs.

The estimates of the mean and variance of the four estimators were used to estimate their bias and mean squared error (MSE), as shown in Table 1. We estimated the standard error of these estimates by repeating the simulation ten times for a "worst case" with $n = 10$ and $t = .0$. We found the standard errors of the bias estimators appear to be less than .003 in all cases, and those for the MSE estimators even smaller (except for the censored variance, which appears to have a standard error on the order of .005). We have therefore reported values in Table 1 to two decimal places.

Note the MLE for the mean is biased, especially for small n. The reason for this is evident when the bias is written in the form $b(\hat{\mu}) = E(c(t)e^{-t^2/2}[S-1] + \overline{X})$, because S is biased low for the standard deviation, 1 (and, of course, $\overline{X}$ has mean 0). The bias of the MLE for the variance is of the form $b(\hat{\sigma}^2) = E([S^2-1]h(t))$, so this estimator is unbiased, consistent with the results shown in Table 1. The MLEs for both moments tend to have slightly smaller variances than have their censored-sample counterparts, as one might expect in part because censoring decreases sample size. Due to bias, however, the estimated MSE of $\hat{\mu}$ is roughly twice that of its censored counterpart (except when P is large, so the censored sample estimate is often equal to the full sample estimate). Both $\hat{\sigma}^2$ and the censored sample esti-

mator appear to be unbiased, but, as expected, MSE($\hat{\sigma}^2$) is significantly smaller than the MSE of the censored sample estimator.

Note that, for larger sample sizes, the estimators exhibit similar performance. This is as expected, since the sample mean and variance are consistent estimators with sampling from any population having a mean and variance. It follows by Equation (6) that the full sample and censored sample estimators of the truncated mean are (weakly) asymptotically equivalent. A similar comment holds for the two estimators of the truncated variance, using Equations (5) and (7).

5. SUMMARY

The mean and variance of a truncated normal population can be calculated easily with common spreadsheet software, using the standard normal and chi-square CDF algorithms implemented there. The mean of a normal population truncated below approaches the truncation point from above, and the variance decreases rapidly to zero, as the truncation point increases. Estimation of these moments can be done easily using the sample moments of corresponding censored samples (when enough data remain), and these estimators are surprisingly good compared to the full sample MLEs. The full-sample estimator of the variance is superior with small sample sizes, because of its relatively small standard error.

[Received September 1995. Revised September 1998.]

REFERENCES

Abramowitz, M., and Stegun, I. (1972), *Handbook of Mathematical Functions*, New York: Dover.

Cohen, A. (1949), "On Estimating the Mean and Standard Deviation of Truncated Normal Distributions," *Journal of the American Statistical Association*, 44, 518–525.

——— (1950), "Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples," *Annals of Mathematical Statistics*, 21, 557–569.

——— (1959), "Simplified Estimation for the Normal Distribution when Samples are Singly Censored or Truncated," *Technometrics*, 1, 217–237.

——— (1961), "Tables for Maximum Likelihood Estimation: Singly Truncated and Singly Censored Samples," *Technometrics*, 3, 433–438.

——— (1991), "Truncated and Censored Samples: Theory and Application," New York: Marcel Dekker.

Fisher, R. (1931), "The Truncated Normal Distribution," *British Association for the Advancement of Science*, 5, xxvi–xxxv.

Gupta, A. (1952), "Estimation of the Mean and Standard Deviation of the Normal Population from a Censored Sample," *Biometrika*, 39, 260–273.

Hald, A. (1952), *Statistical Theory with Engineering Applications*, New York: Wiley.

Halperin, M. (1952), "Estimation in the Truncated Normal Distribution," *Journal of the American Statistical Association*, 47, 457–465.

Johnson, N., and Kotz, S. (1970), *Continuous Univariate Distributions—1*, New York: Wiley.

Kececioglu, D. (1991), *Reliability Engineering Handbook* (vol. 1), Englewood Cliffs: Prentice-Hall.

Law, A., and Kelton, W. (1991), *Simulation Modeling and Analysis*, New York: McGraw-Hill.